UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

AUTHORS GUILD, DAVID BALDACCI, MARY BLY, MICHAEL CONNELLY, SYLVIA DAY, JONATHAN FRANZEN, JOHN GRISHAM, ELIN HILDERBRAND, CHRISTINA BAKER KLINE, MAYA SHANBHAG LANG, VICTOR LAVALLE, GEORGE R.R. MARTIN, JODI PICOULT, DOUGLAS PRESTON, ROXANA ROBINSON, GEORGE SAUNDERS, SCOTT TUROW, and RACHEL VAIL, individually and on behalf of others similarly situated,	Case No. 1:23-cv-08292-SHS Case No. 1:23-cv-10211-SHS Case No. 1:24-cv-00084-SHS **STIPULATION & ORDER RE: DISCOVERY OF ELECTRONICALLY STORED INFORMATION**

Plaintiffs,

v.

OPENAI INC., OPENAI OPCO LLC, OPENAI GP LLC, OPENAI LLC, OPENAI GLOBAL LLC, OAI CORPORATION LLC, OPENAI HOLDINGS LLC, OPENAI STARTUP FUND I LP, OPENAI STARTUP FUND GP I LLC, OPENAI STARTUP FUND MANAGEMENT LLC, and MICROSOFT CORPORATION,

Defendants.

JONATHAN ALTER, KAI BIRD, TAYLOR BRANCH, RICH COHEN, EUGENE LINDEN, DANIEL OKRENT, JULIAN SANCTON, HAMPTON SIDES, STACY SCHIFF, JAMES SHAPIRO, JIA TOLENTINO, and SIMON WINCHESTER, on behalf of themselves and all others similarly situated,

Plaintiffs,

v.

OPENAI, INC., OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, OPENAI HOLDINGS, LLC, and MICROSOFT CORPORATION,

Defendants.

[Caption continued on next page]

1

NICHOLAS A. BASBANES and NICHOLAS
NGAGOYEANES (professionally known as
Nicholas Gage), individually and on behalf of all
others similarly situated,

                    Plaintiffs,

          v.

MICROSOFT CORPORATION, OPENAI, INC.,
OPENAI GP, L.L.C., OPENAI HOLDINGS, LLC,
OAI CORPORATION, LLC, OPENAI GLOBAL,
LLC, OPENAI, L.L.C., and OPENAI OPCO, LLC,

                    Defendants.

To expedite the flow of discovery material and to facilitate the consistency in the format of the documents to be produced by the Parties in this case, Plaintiffs and Defendants (collectively "Parties"), by and through their respective counsel, hereby stipulate and agree to the terms of this [Proposed] Stipulation & Order Re: Discovery of Electronically Stored Information (the "ESI Protocol" or "Protocol").

1.   This Order supplements all other discovery rules and orders. It streamlines Electronically Stored Information ("ESI") production to promote a "just, speedy, and inexpensive determination" of this action, as required by Federal Rule of Civil Procedure 1.

2.   This Order may be modified in the Court's discretion or by stipulation.

3.   The Parties are aware of the importance the Court places on cooperation and commit to cooperate in good faith throughout the matter consistent with this Court's expectations.

4.   The Producing Party will conduct a reasonable investigation to identify custodians and use search terms reasonably calculated to identify responsive materials consistent with the

parties' discovery obligations pursuant to the Federal Rules of Civil Procedure. The Producing Party will in the first instance propose search terms and custodians and corresponding hit reports. The parties will negotiate in good faith, including by meeting and conferring, to identify custodians and search terms. The Producing Party will provide hit reports for proposed search terms, including search terms proposed by a Requesting Party after the date of this Order, to inform such negotiations. If the Requesting Party makes a proposal for additional search terms for current custodians, the Producing Party must respond within five business days by either accepting a proposal, rejecting it and providing a hit report for the Requesting Party's proposal, or making a counterproposal and providing a hit report for both the Requesting Party's proposal and the Producing Party's counterproposal. If the Requesting Party makes a proposal for additional custodians, the Producing Party must respond within ten business days by either accepting a proposal, rejecting it, or making a counterproposal and providing a hit report.

5.   The parties will disclose at the outset any use of enterprise messaging tools, such as Slack or MS Teams, including the type of subscription and retention settings in place. The parties will meet and confer regarding issues relating to the parameters for collection of such data, the format of the production, and other unique issues to this type of data.

6.   Nothing in this Order shall be construed to waive a requesting party's right to move the Court to compel production of documents from additional custodians, nor shall anything in this Order be construed to waive a producing party's right to oppose any such motion to compel.

7.   Nothing in this Order prevents the parties from agreeing to use technology assisted review and other techniques insofar as their use improves the efficacy of discovery.

8.   If a party elects to use Technology Assisted Review (TAR),[1] the parties will meet and confer in good faith regarding the use of TAR and proposed TAR Protocol.

9.   Production Format for ESI.

   a.   Except as otherwise provided in this Order, all documents existing in electronic format (e.g., Word, PDF, webpage files) shall be produced as 300 DPI TIFF black and white images with 1-bit depth and with a Bates number stamped on each page, document metadata (including extracted or OCR text), and a link to the associated extracted or OCR text file.   The parties will consider re-producing select and specifically identified color documents as paginated color JPGs upon reasonable request by the receiving party.

   b.   Source code and certain other highly confidential technical materials will be produced in native format and shall be made available for inspection on a secure computer in a secured room without Internet access or network access to other computers at a secure location determined by the producing party, pursuant to the terms of the Protective Order governing "HIGHLY CONFIDENTIAL – SOURCE CODE" materials.

   c.   Spreadsheet-type files (e.g., Microsoft Excel), PowerPoint files, and Multimedia files (video, audio, voicemail, etc.)_should be produced in native format unless they require redaction. Documents produced in native format should be named according to the Bates number assigned, with the Confidentiality designation appended to the Bates numbers with a hyphen. Single-page Bates-stamped TIFF image slip-sheets will be included for each document produced in native format. The slip-sheets will display the Bates number of the native file, the Confidentiality endorsement, and an

---

[1] TAR refers to the application of machine learning and other advanced analytical technologies to the process of identifying, coding, and categorizing ESI based on relevance to specific legal matters. TAR methodologies may include predictive coding, clustering, and conceptual analysis. For the avoidance of doubt, TAR does not include the use of keyword searches.

endorsement stating, "File Produced Natively."

d. For Excel and PowerPoint type files that are printed to TIFF for redaction and redacted, the following printing options shall be enabled:

    i. Excel Print to TIFF Options

1. Unhide columns and rows
2. Unhide worksheets
3. Autofit columns and rows, settings to be over by columns first and, then down by rows
4. Print gridlines
5. Do not apply Autofilter
6. Display headings
7. Display comments
8. Header and Footer filename field handling: Show field code

    ii. PowerPoint Print to TIFF Options

1. Print notes pages
2. Print hidden slides
3. Print comments

e. The Producing Party shall also make reasonable efforts to ensure that any spreadsheets produced only as TIFF images because of redactions are formatted so as to be legible. For redacted items which were originally ESI, all metadata fields noted in this protocol that do not contain privileged or protected information will be provided and will include all non- redacted data. A Document's status as redacted does not relieve the Producing Party from providing all of the metadata required herein.

f. All attachments, addendums,enclosures, and/or exhibits to a parent document will be, to the extent reasonably possible, produced immediately following the parent document and identified as they relate to the respective parent document with the parent/child relationship intact.

g. The parties are in the process of negotiating a protocol regarding the

inspection of training data (the "Training Protocol"). To the extent the

provisions of the Training Protocol conflict with this Order, the provisions

of the Training Protocol shall control as to the inspection of training data.

10.   Metadata. Load or Concordance DAT files should include, where applicable and

to the extent it exists and was captured at the time of the collection, the information listed in

Exhibit A attached. To the extent that metadata does not exist or is not reasonably accessible or

available for any documents produced, nothing in this Order shall require any party to extract,

capture, collect or produce such metadata.

11.   Privilege Logs. A party that withholds documents, ESI, and any other materials

and information on the grounds of attorney-client privilege, attorney work product, and/or any

other applicable privilege, immunity or protection from discovery, shall produce a privilege log.

The following documents and communications need not be included on a privilege log: (1)

privileged communications exclusively between a party and its outside or in-house counsel on or

after September 09, 2023; (2) work product performed by a party at the direction of its outside or

in-house counsel on or after September 09, 2023; and (3) privileged or work product documents

or communications between a party and its outside or in-house counsel related to the following

lawsuits and investigations: *Doe 1, et al. v. GitHub, Inc. et al.*, Case No. 4:22-cv-06823-JST

(N.D. Cal.); *In re OpenAI ChatGPT Litigation*, Case No. 3:23-cv-03223-AMO (N.D. Cal.);

Authors Guild, et al., v. OpenAI, Inc, et al., Case No. 1:23-cv-08292-SHS (S.D.N.Y.); *Alter, et*

al. v. OpenAI Inc. et al., 23-cv-10211-SHS (S.D.N.Y.); *The New York Times Company v.*

Microsoft Corp. et al., 23-cv-11195-SHS (S.D.N.Y.); *Basbanes et al. v. Microsoft Corp. et al.*,

24-cv-00084-SHS (S.D.N.Y.); *The Intercept Media, Inc. v. OpenAI, Inc. et al.*, Case No. 1:24-

cv-01515-JSR (S.D.N.Y.); *Raw Story Media, Inc. et al v. OpenAI Inc. et al.*, 1:24-cv-01514-CM

(S.D.N.Y.); the pending Federal Trade Commission Investigation, FTC File No. 232-3044, and

any other subsequently filed or commenced lawsuits or investigations relating to OpenAI. The

parties will meet and confer to negotiate a reasonable date for the exchange of privilege logs.

The privilege log shall be produced in Excel format and include:

a. A unique number for each entry on the log.

b. Date of document for all ESI and to the extent known or discernible for all hard copy documents. For emails this should be the sent date of the document and for loose ESI this should be the last-modified or create date of the document.

c. Custodian of the document. For emails this should be populated with the metadata extracted from the Custodian(s) field. For loose ESI, this should be populated with the Custodian value assigned to the instance of the document in question.

d. Author of the document. For emails this should be populated with the metadata extracted from the "Email From" field associated with the file. For loose ESI, this should be populated with the metadata extracted from the "Author" field. For hard copy documents, this will be populated with the "Custodian."

e. Recipient(s) of the document where reasonably ascertainable. For emails this will be populated with the metadata extracted from the "Email To" field associated with the file. Separate columns will be included for the metadata extracted from the "Email CC" and "Email BCC" fields, where populated.

f. The type of privilege being asserted, for example: (a) AC for Attorney/Client, (b) WP for Attorney Work Product, (c) CI for Common Interest.

g. The reason for the redaction. If more than one reason, separate by semi-colons.

h. Nothing in Paragraph 15 of this Order shall be read to preclude the parties from logging documents categorically or read to be inconsistent with Local Rule 26.2.

12. Any Party that issues a non-party subpoena shall timely notify other Parties when

it receives non-party productions of documents, ESI or other responses from the non-party, and shall provide copies of all such productions in the format in which they were received from the non-party to the other Parties. If the non-party produces documents or ESI that are not Bates labeled, the Receiving Party will Bates-label the documents or ESI before producing a copy to the other Parties if the produced documents or ESI are of the type so requiring consistent with other provisions of this Order. Nothing in this Protocol precludes a Receiving Party from reviewing a non-party production of documents or ESI and applying (if the non-party produced documents without a confidentiality designation) or elevating a confidentiality designation (e.g., elevating documents designated as CONFIDENTIAL to ATTORNEYS' EYES ONLY) if the production contains confidential or extremely sensitive information.

13.   Technical Variances. Recognizing that each Producing Party may experience production issues due to data systems or files that may not be fully compatible with the technical specifications set forth herein, any practice or procedure set forth herein may be varied by agreement of the Parties, confirmed in writing, to accommodate such issues and/or where such variance is deemed appropriate to facilitate the timely and economical production of Documents or ESI. No Party shall unreasonably object to any such variance.

14.   Reproduction as Natives. The Parties agree that to the extent any Party seeks production in native format of specifically identified ESI produced originally in TIFF form, the Producing Party shall respond reasonably and in good faith to any such request.

15.   Pursuant to Fed. R. Evid. 502(d) and (e), the disclosure during discovery of any communication, information, or material that is protected by the attorney-client privilege or work-product protection, as defined by Fed. R. Evid. 502(g), shall not waive the privilege or protection in the above-captioned case, or any other federal or state proceeding, for either that communication, information, or material or the subject matter of the same, unless there is an intentional waiver of the privilege or protection to support an affirmative use of such communication, information, or material in support of the Party's claim or defense, in which event the scope of any such waiver shall be determined by Fed. R. Evid. 502(a)(2) and (3).

16.   In the event of an inadvertent production of material protected by the attorney-client privilege or work-product protection, the producing party may notify the receiving party within 30 days after the producing party becomes aware of such inadvertent production. Upon notification, the receiving party shall immediately return, sequester, or destroy the inadvertently-produced Discovery Material and all copies as well as notes, summaries, and/or work product reflecting the content of such material. If only a portion of the Discovery Material is privileged, the producing party shall immediately provide a new version of the Discovery Material in which the privileged material has been redacted. No further use or disclosure shall be made of the inadvertently produced Discovery Material, and the receiving party shall take all reasonable and appropriate steps to retrieve and destroy the Discovery Material, and all copies, from any person who has received the Discovery Material through the receiving party.

        IT IS SO STIPULATED, through Counsel of Record.

Dated: Sep. 11, 2024   /s/ Alejandra Salinas
                                          Counsel for Plaintiff

Dated: Sep. 11, 2024   /s/ Christopher S. Sun
                                          Counsel for Defendant

Dated: Sep. 11, 2024   /s/ Jared B. Briant
                                          Counsel for Defendant

        IT IS ORDERED that the forgoing Agreement is approved.

Dated: September 13, 2024

                                          ONA T. WANG
                                          United States Magistrate Judge

9

EXHIBIT A

Field	Data Type	Paper	Native Files & Email Attachments	Email
BegDoc	Integer - Text	Starting Bates #	Starting Bates #	Starting Bates #
EndDoc	Integer - Text	Ending Bates #	Ending Bates #	Ending Bates #
BegAttach	Integer - Text	Starting bates # of document family	Starting bates # of document family	Starting bates # of document family
EndAttach	Integer - Text	Ending bates # of document family	Ending bates # of document family	Ending bates # of document family
Custodians	Text – paragraph Separate entries with ";"		All names of people the document was collected from even if removed from production as a duplicate	All names of people the document was collected from even if removed from production as a duplicate
From	Text - paragraph			Sender of message
To	Text – paragraph Separate entries with ";"			Recipients of message
CC	Text – paragraph Separate entries with ";"			Copied recipients
BCC	Text – paragraph Separate entries with ";"			Blind copied recipients
Subject	Text - paragraph			Subject of message
DateSent	Date (mm/dd/yyyy)			Date message sent
DateReceived	Date (mm/dd/yyyy)			Date message received

Field	Data Type	Paper	Native Files & Email Attachments	Email
FileName	Text - paragraph		Name of original file including extension	Name of original file including extension
FileExtension	Text		Extension of original file	Extension of original file
DateCreated	Date/Time (mm/dd/yy yy)		Date file was created	
DateModified	Date/Time (mm/dd/yy yy)		Date file was last modified	
Author	Text - paragraph		Document author from metadata	
Confidentiality	Single choice	Designation branded onto the document	Designation branded onto the document	Designation branded onto the document
FullText	Link	Link to the extracted or OCR text file	Link to the extracted or OCR text file	Link to the extracted or OCR text file
DocLink	Link	Link to the native file	Link to the native file	Link to the native file
MD5Hash	Text		The MD5Hash value of the file	The MD5Hash value of the file
AttachNam	Text – paragraph Separate entries with ";"			File name of the attachment, with any attachments separated by semi-colon.
PgCount	Integer	Number of pages in the item	Number of pages in the item	Number of pages in the item
Custodian	Text – paragraph	Name of person or source from whose/which files the item is produced	Name of person or source from whose/which files the item is produced	Name of person or source from whose/which files the item is produced
Parent Date/Sort Date	Date/Time (mm/dd/yy yy)		The date associated with a family's parent record, which is assigned as follows:	The date associated with a family's parent record, which is assigned as follows:

Field	Data Type	Paper	Native Files & Email Attachments	Email
			Loose Edocs populated with 1st occurrence of Date Last Modified or Date Created.	Emails populated with 1st occurrence: Date Sent or Date Received or Date Last Modified. Email attachments populated from the date above. Outlook Appointments populated with 1st occurrence of Start Date or Date Last Modified.
TimeReceived	Time			Time email or calendar item was received (Date and time fields may be combined)
FileSize	Integer		Size (in kilobytes) of the source native file	Size (in kilobytes) of the source native file
SourceFilePath	Text – paragraph		The directory structure or path where the original file was stored on the Party's source computer system, ending in the filename. Any container name (such as ZIP or PST containers) is included in the path.	The directory structure or path where the original file was stored on the Party's source computer system, ending in the filename. Any container name (such as ZIP or PST containers) is included in the path.
SourceParty	Text – paragraph	Name of entity or Party producing the item	Name of entity or Party producing the item	Name of entity or Party producing the item
RecordType	Single choice		Indicates item type (e.g., email, edoc, attachment)	Indicates item type (e.g., email, edoc, attachment)

Field	Data Type	Paper	Native Files & Email Attachments	Email
FileType	Single choice		(e.g., Outlook, Adobe Acrobat, MS Word, etc.)	(e.g., Outlook, Adobe Acrobat, MS Word, etc.)
TimeSent (hh:mmAM/PM)	Time			(e.g., Outlook, Adobe Acrobat, MS Word, etc.)
Number of Attachments	Integer			Number of attached, embedded or grouped items
TimeCreated (hh:mm AM/PM)	Time		Time the item was created (Date and time fields may be combined)	
ModifiedBy	Text – paragraph		Person who last modified or saved the item, as populated in the metadata or document properties of the native file	
LastModTime (hh:mm AM/PM)	Time		Time the item was last modified	
Title	Text - paragraphe		Any value populated in the Title field of the source file metadata or item properties	
Redacted	Single choice	User-generated field that will indicate redactions. With the word "REDACTED". Otherwise, blank.	User-generated field that will indicate redactions. With the word "REDACTED". Otherwise, blank.	User-generated field that will indicate redactions. With the word "REDACTED". Otherwise, blank.

14